UNCERTAINTY STOCKTAKE – SCOPING PHASE

Discussion Paper to TASM Division, Department for Transport

Richard Batley, ITS Leeds

5th July 2018

EXECUTIVE SUMMARY AND RECOMMENDATIONS

Current state of TAG guidance on uncertainty

WebTAG encompasses guidance on various aspects of 'uncertainty' (broadly defined) within the modelling and appraisal process, and the quality of some aspects of this guidance (e.g. on scheme costs) is very respectable. Furthermore, the ongoing TASM work on scenario analysis might be considered best (even leading) practice for national transport ministries. However, in totality, TAG guidance on uncertainty gives the impression of having evolved in a rather ad hoc and fragmented fashion over time.

Recommended actions for strengthening TAG guidance on uncertainty

R1: The priority should be to review – and where necessary refresh – existing WebTAG assets, with the objective of delivering a coherent 'toolkit' to deal with uncertainty in transport modelling and appraisal.

R2: The purpose of the 'toolkit' should be to equip practitioners with sufficient guidance to enable them to investigate and report key sources of uncertainty in modelling and appraisal in a systematic and consistent fashion. To these ends, the guidance could usefully encompass the following areas:

i) Identify sources of input uncertainty (encompassing any errors in data, assumptions concerning future planning variables and any assumed parameters) and model error (encompassing any error in estimated parameters and any specification error) on both demand and supply sides.

ii) Where appropriate, outline acceptable ranges for assumed parameters. Otherwise, issue guidance on reporting expectations with regards to observed variances in data inputs and estimated parameters.

iii) Outline methods for quantifying key sources of uncertainty both individually and collectively, and especially the propagation of uncertainty through model and appraisal systems.

iv) Discuss approaches to mitigating these uncertainties.

v) Issue guidance on reporting expectations with regards to overall uncertainty in modelling and appraisal forecasts.

R3: As and when such guidance is developed and released, a longer term priority (informed by item iii) above) should be to form an overview on the key sources of uncertainty – especially if these are in areas where the Department could take action to mitigate (e.g. commission new research on fuel cost elasticities, commission new data collection on VTTS, etc.).

R4: A three-package work programme is proposed for drafting, testing and finalising the proposed uncertainty 'toolkit', as follows:

- WP1: Draft guidance on the uncertainty 'toolkit'
- WP2: Case study analysis of uncertainty mitigation
- WP3: Update guidance on the uncertainty 'toolkit' and issue research recommendations

R5: The optimal team for this work programme would be some combination of practitioner (bringing insight on practical context) and academic (bringing insight on methods), but would need active contribution from the Department.

1. INTRODUCTION

This note reports a 'stocktake' commissioned by the Transport Appraisal and Strategic Modelling (TASM) Division of the Department for Transport, to assist them in shaping their requirements in the area of 'uncertainty in modelling and appraisal'.

The specific objectives of the stocktake were three-fold, namely:

1. To deliver a 'think-piece' to inform TASM thinking on possible updates to WebTAG guidance on the treatment of uncertainty. This work would help to inform the specification of more detailed research which might be undertaken, and in particular the focus of research to consider variability around key inputs and model parameters which would be the focus of an uncertainty stocktake. Gather intelligence about different areas of uncertainty and methodologies for addressing uncertainty in modelling and appraisal;

2. Create an indicative list of research priorities; and

3. Provide suggestions for the possible format of the research.

A budget of 5 days was agreed to undertake this task.

The layout of the note is as follows. The following section (Section 2) outlines the background to the stocktake exercise. Section 3 introduces definitions and concepts. Section 4 summarises some key contributions to the academic literature on uncertainty in transport modelling and appraisal. Section 5 documents the areas of WebTAG which include representation of uncertainty (and related concepts), whilst Section 6 highlights presentational devices which have previously been used by DfT and similar agencies to communicate aspects of uncertainty in modelling and appraisal.

Having carefully set out the context of the discussion – from both conceptual and practical angles – through Sections 2-6, this note culminates in Section 7, which addresses several specific questions designed to help the Department determine its research priorities in the area of uncertainty and what actions might be taken to address these priorities.

2. BACKGROUND TO THE STOCKTAKE

The notion of uncertainty in transport modelling and appraisal is complex and multi-faceted. It would seem to mean different things to different agents within the modelling, appraisal (and policy) process. As of consequence, any single agent in the process (e.g. modeller, appraisal analyst, DfT official) may not necessarily have full oversight of the inter-relationships between different sources of uncertainty (broadly defined), and/or their cumulative effect on the precision and robustness of appraisal outcomes.

In what follows, this document will seek to introduce these different facets of uncertainty, indicate the manner in which they might interact, and identify some key areas of focus for TASM's future research programme. However, before proceeding to such detail, it may be useful to introduce two broad dimensions of uncertainty which motivate current policy interest in this area.

The first dimension arises from a recognition that, in practical modelling and appraisal, there are shortcomings in systematically considering the relationship between the precision of models (in terms of both inputs and outputs) and the robustness of appraisal results. As a case in point, this issue has recently come under scrutiny in the context of the business case for the proposed High Speed 2 (HS2) rail scheme in the UK, where the National Audit Office (NAO) recommended that: *'The Department* (for Transport) *and HS2 should recognise explicitly the uncertainty in the economic case by quoting ranges rather than a point estimate. The risks and uncertainty to the benefit-cost ratio have not been clearly stated'* (NAO, 2013 p12).

Partly in response to such critiques, the Department has developed an 'Analytical Assurance Framework' (DfT, 2013), which endeavours to promote: *'the correct balance between robustness, timeliness, and cost, for the decision at hand'* (p4). The Framework encompasses three key dimensions, namely: 1) the potential for challenge to the analysis; 2) the risks of an error in the analysis and the uncertainty inherent in the analytical advice to the Permanent Secretary and Secretary of State; and 3) the degree to which error and/or uncertainty has been reduced.

The second dimension arises from a recognition that practical modelling and appraisal is being conducted over longer timeframes (60 years for significant assets) within an increasingly uncertain world – subject to uncertainties not only in behaviour (on both demand and supply sides), but also exogenous uncertainties (such as the natural environment, the economy, as well as political and regulatory change). Recognising such uncertainties, the NIC's (2017) recent 5-year assessment of the UK's infrastructure requirements modelled a range of scenarios across the solid waste, water, transport and energy sectors in order to explore different future pressures on infrastructure. NIC noted: '[These] *scenarios are constructed using the four drivers and provide contrasting versions of the future. They will be used as reference points against which to sense-check infrastructure options, decisions and recommendations and provide a robust way of taking into account the substantive uncertainty when looking out to 2050'*.

Drawing together these two dimensions of uncertainty, the all-day workshop of the Joint Analysis Development Panel (JADP) held in November 2017 identified 'uncertainty' (broadly defined) as one of the areas where DfT should prioritise its future research. On the back of that workshop, the Department commissioned this think-piece to highlight and prioritise the areas where WebTAG could be updated to better reflect uncertainty.

In the scope for the think-piece, DfT noted that:

'WebTAG already covers a number of aspects of uncertainty as it relates to scheme appraisal. These areas include:

- Sensitivity tests in forecasting (such as those defined in A5.3 Rail Appraisal);

- Uncertainty Log and national uncertainty range (M4 Forecasting and Uncertainty);

- Uncertainty in appraisal (ranges around VTTS estimates); and

- Issues related to uncertainty in long-term demand & benefits (A5.3 and A1.1).

The Department is also continuing to develop the use of scenario analysis. The last published Road Traffic Forecasts (RTF15) contained forecasts under five different scenarios. We are currently developing new traffic scenarios to support the development of the Second Road Investment Strategy (RIS2) and a future Road Traffic Forecast publication. In parallel we are working on developing a common set of DfT scenarios, which could be used consistently across the Department in the appraisal of strategic programmes or schemes where required. Separately to this, bespoke scenarios have been used in aviation forecasts (looking into aviation-specific uncertainties) and a bespoke uncertainty analysis was undertaken in HS2 business cases (Monte Carlo analysis).

Recent relevant publications include updated guidance on the Valuation of Travel Time Savings and associated uncertainty ranges around the estimates (published 2016) and research into long-term benefits of transport investments alongside changes to unit A5.3 and A1.1 of WebTAG guidance'.

3. SOME DEFINITIONS AND CONCEPTS

According to the Oxford English Dictionary, **uncertainty** is defined as *'The quality of being uncertain in respect of duration, continuance, occurrence, etc.; liability to chance or accident. Also, the quality of being indeterminate as to magnitude or value; the amount of variation in a numerical result that is consistent with observation.'*. Cross-referencing Norvig & Thrun (2018), Wikipedia defined **uncertainty** as: *'…a situation involving ambiguous and/or unknown information. It applies to predictions of future events, to physical measurements that are already made, or to the unknown. Uncertainty arises in partially observable and/or stochastic environments, as well as due to ignorance, indolence, or both'.*

Wikipedia further noted that specialists in decision theory, statistics and other quantitative fields have defined **uncertainty**: *'the lack of certainty, a state of limited knowledge where it is impossible to exactly describe the existing state, a future outcome, or more than one possible outcome'*; the **measurement of uncertainty**: *'a set of possible states or outcomes where probabilities are assigned to each possible state or outcome – this also includes the application of a probability density function to continuous variables'*; and **second order uncertainty**: *'in statistics and economics, second-order uncertainty is represented in probability density functions over (first-order) probabilities'*.

Turning specifically to economics, which typically entails a behavioural dimension, some authors, notably Keynes (1921, 1936) and Knight (1921) have distinguished **risk** from **uncertainty**, where the former refers to situations where the probability of an outcome/state of nature is known to an individual decision-maker, and the latter where probability is unknown.

Any practical transport model or appraisal is likely to give rise to multiple sources of uncertainty. In this context, it is not simply the prevalence and magnitude of these sources of uncertainty that is of interest, but also their interaction and propagation through the modelling/appraisal system. Wikipedia defined the **propagation of uncertainty** as *'the effect of variables' uncertainties on the uncertainty of a function based on them'*. In the specific context of transport modelling, Rasouli & Timmermans (2012) defined **propagation** as the case where *'…the ultimate forecast involves a series of successive sub-models in which the output of the previous sub-model in the model chain is used as input to the next sub-model,* [and where] *errors in any sub-model may be amplified or reduced in the next sub-models'*.

Finally, it might be noted that statisticians often use the term **error** (or, more precisely, random error) synonymously with uncertainty, and this convention will be obeyed in what follows. If **absolute error** in a variable is given by $\Delta x = x - \bar{x}$ then **relative** (or percentage) error is given by $\Delta x = x/\bar{x}$. In many applications, error is represented in terms of the **standard deviation** $\sigma = \sqrt{\sum_n (\Delta x_n)^2 / n - 1}$ thereby allowing the variable and the associated error to be expressed as an interval $\bar{x} \pm u$. Imposing more specific distributional assumptions, if $x \sim N$, then $\bar{x} \pm \sigma$ will cover the true value of x in around 70% of cases. The above assumes that errors are uncorrelated, and if this does not hold then covariance must also be taken into account.

To promote a degree of simplification, this note will generally adhere to the Department's terminology of 'uncertainty' – but interpreted in its broadest sense to encompass all of the above concepts. Where appropriate, the note will however distinguish between the specific concepts detailed above.

4. RAPID LITERATURE REVIEW

There is an extant literature – albeit somewhat limited in size and scope – on the representation of uncertainty within transport modelling and appraisal, especially in relation to travel demand forecasting. This section summarises some of the key contributions to this literature, but should be seen as an expedient (rather than a complete and definitive) review of the literature.

4.1 Pearman (1976)

Motivating his paper, Pearman remarked that techniques for the appraisal of transportation investment projects in the presence of uncertainty *'have been frequently and justifiably criticised'*. He identified three 'conventional' approaches to such appraisal:

a) First, he referred to techniques which react to the presence of uncertainty by adopting *'an essentially conservative standpoint'*. For example, in the case of a road scheme, this might entail pessimistic forecasts of land acquisition costs, accident costs, time savings, etc. Furthermore, conservatism might extend to forecasts of the economic life of the scheme, or the inclusion of a 'risk premium' representing uncertainties in the flow of benefits.
b) Second, and with reference to Keynes' and Knight's dichotomy noted in Section 3 above, he referred to techniques for re-conceptualising uncertainty as risk, through objective estimation of the probability of occurrence of each possible state of nature. Once this has been done, the

appraisal can then be progressed in terms of maximising the expected present value of the stream of net benefits.

c) Third, he referred to techniques which deal with the situation of *'complete ignorance'* where there is no knowledge of the probability of each state of nature. These techniques encompass the likes of maximin, maximax, the Hurwicz α criterion, and maximax, which have attracted interest in the theoretical literature, but seen limited application in practice.

Having reviewed these approaches, the paper proceeds to develop a 'new' method for handling uncertainty in investment appraisal. This method takes advantage of the subjective judgements of the decision-maker and available forecasts of a more objective nature, to develop an a priori ranking of the likelihood of future states of nature. Pearman noted that the method is appropriate in situations where results may reasonably be averaged over a series of investments. It is claimed that the method is simple to apply, and more reliable than conventional techniques of the decision theoretic literature, such as maximin, maximax, etc. A simple numerical illustration is given in the paper.

4.2 Ashley (1980)

In the course of a Department of Transport (DoT) sponsored study commissioned in the wake of the Leitch Report (1977), Ashley (1980) developed a framework for analysing model uncertainty. In particular, Ashley's study responded to one of Leith's recommendations, namely that:

'The Department should indicate the likely range of uncertainties involved in the forecasts and demonstrate the consequences of selecting different values within the likely range. It should never put itself in the position of appearing to defend a single figure as if it were uniquely correct'.

In this context, model uncertainty is interpreted in terms of 'model error', and Ashley laid out three principles governing the analysis of this error:

a) The first principle of model error analysis is that there are a number of distinct sources of error, each with its own special characteristics that must be individually determined.
b) The second principle concerns the approach to calculating these errors. Because of lack of research, we must fall back on subjective reasoning rather than objective analysis.
c) The third principle relates to the interaction of the various sources of model error.

He further dichotomised between two basic classes of 'forecast error' in transport models:

1) The accuracy of the forecast exogenous input variables (generally planning data and economic parameters.
2) The accuracy of each of the individual sub-models (highway traffic model, car ownership model, trip end model, distribution and assignment models).

Within 2), it is postulated that model error arises from two sources:

2a) Calibration (or estimation) error

If we have confidence in the specification of the model, then calibration error can be determined by statistical procedures. That is to say, if there is a known and exact relationship then our ability to estimate parameters will be limited only by the quality of the data (and especially the volume of the data).

2b) Specification error

This error will appear to be small at calibration, since the process of model selection typically ensures that observed data is reproduced well. The extent of any such error is more likely to become apparent in the course of forecasting.

Ashley noted that the main inputs to transport models are forecasts of planning data such as population and employment, and forecasts of economic parameters such as household income and fuel costs. It is suggested that, from comparison of model forecasts of planning data growth over say a 20-year period with actuals, a standard error of the growth forecasts can in principle be calculated, the PDF for which appears approximately Normal. The question then arises as to whether the planner can improve on this forecast by using local data.

Ashley suggested that a similar approach can be followed for economic variables and income in particular. Historical data on GDP is available from ONS, making it possible to develop GDP forecasting models which extrapolate into the future based on past trends[1]. In the case of fuel costs, by contrast, it is noted that future trends will have little relationship to past trends. Therefore, it makes sense to defer to the Department's official fuel costs. Ashley pointed out that, if the error within income forecasts is increased to reflect the effect of oil prices on the economy, then this implies that fuel prices and GDP are in some way correlated – which might be considered controversial.

Turning to the issue of error propagation, Ashley stated the following general model (which is essentially a Taylor series expansion):

$y = f(x_1 + x_2, \dots, x_n)$

$e_y^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial f}{\partial x_i} \cdot \frac{\partial f}{\partial x_j} . r_{ij} . e_{x_i} . e_{x_j}$ where:

e_y and e_{x_i} are the (standard) errors of y and x_i

r_{ij} is the correlation between x_i and x_j

$\frac{\partial f}{\partial x_i}$ is the partial derivative of f with respect to x_i

[1] Note that HM Treasury publishes forecasts for the current year at https://www.gov.uk/government/collections/data-forecasts#2018

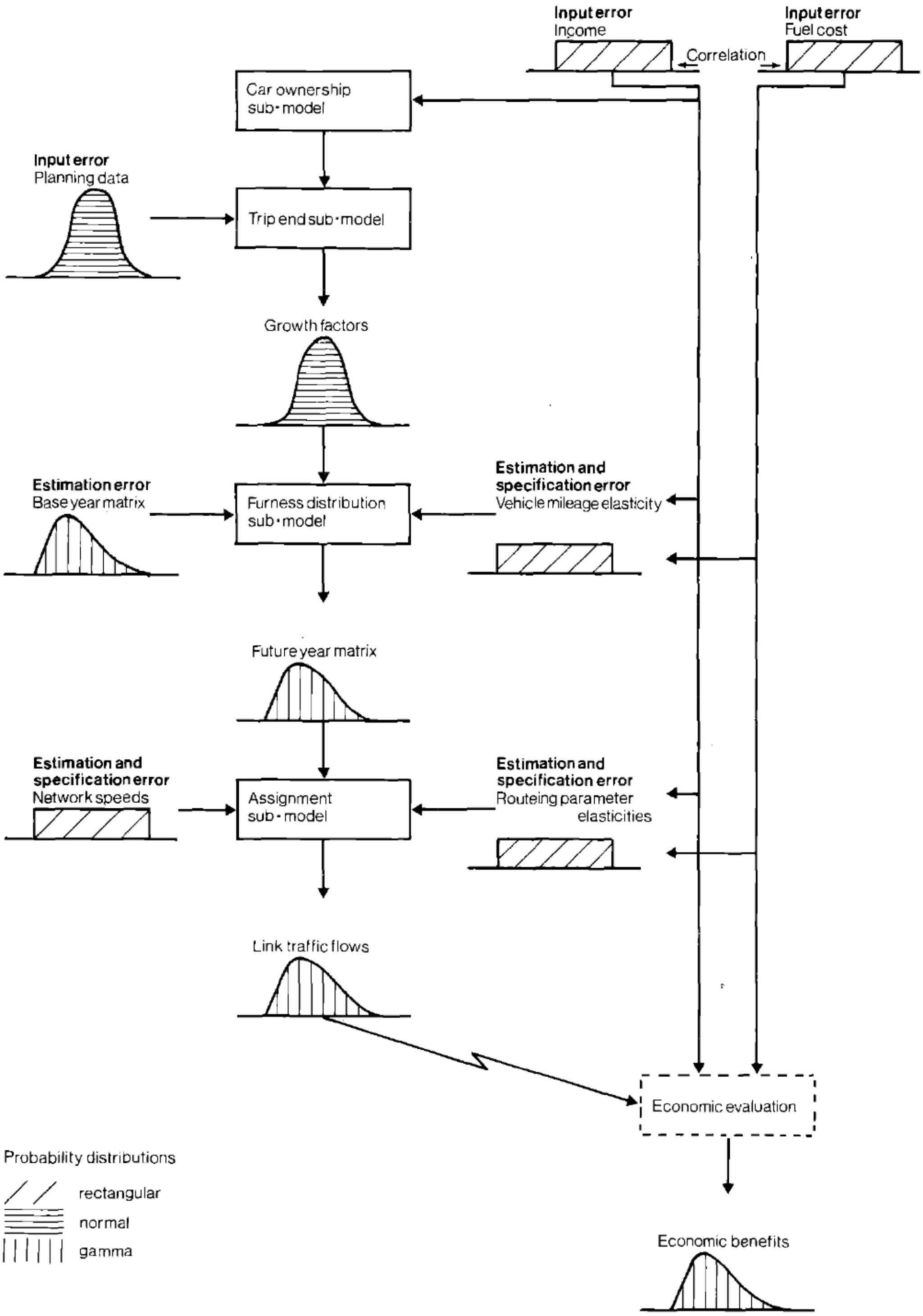

Fig. 2. The highway traffic model illustrating the main sources of error.

(Source: Ashley (1980))

Despite the relative clarity of this theoretical approach, Ashley noted the practical difficulty of deriving the partial derivatives, due to the complex inter-relationships and iterative techniques inherent in the likes of distribution and assignment sub-models. As a pragmatic alternative, Ashley deployed an error simulation approach – summarised in the following figure for the example of a traditional four-stage transport model.

The key features of this approach are as follows:

- Car ownership sub-model: this is subject to input error in relation to incomes (which is assumed to have a rectangular distribution).
- Trip end sub-model: this is subject to error in relation to planning data (Normal).
- Distribution sub-model: this is subject to input error in relation to growth factors (Normal), estimation error in relation to the base year matrix (gamma), and estimation and specification error in relation to the vehicle mileage elasticity (rectangular).
- Assignment sub-model: this is subject to input error in relation to the future year matrix (gamma), estimation and specification error in relation to both network speeds (rectangular) and the routing parameter elasticity (rectangular).
- Economic evaluation: this is subject to input errors in relation to incomes (rectangular), fuel costs (rectangular), and link traffic flows (gamma).

4.3 Zhao & Kockelman (2002)

Zhao & Kockelman's paper was motivated by the observation that transport models entail considerable uncertainty – in both inputs and model parameters. The paper acknowledges that, whilst modellers can do relatively little about errors due to mis-measurement, poor sampling, mis-computation, model mis-specification, and data aggregation, purely stochastic errors can be accommodated statistically and explicitly. Three sources of these stochastic errors are highlighted, namely:

a) inherent uncertainty
b) input uncertainty
c) propagated uncertainty

Since travel demand model parameters are random variables, estimated from samples of the population, model estimates are associated with variations and co-variations. These variations constitute inherent uncertainty. Also, the use of predictions of future demographic data (e.g., employment and land use) as inputs to traffic demand forecasting models contributes input uncertainty. Moreover, since transport demand models are generally estimated and applied sequentially, the results or estimates of one model act as input to subsequent models. Therefore, uncertainty is passed forward through the modelling system, giving rise to propagated uncertainty. The cumulative impact of these three forms of uncertainty is the focus of this paper.

Zhao & Kockelman investigated this issue by quantifying the variability in model inputs, such as zonal socio-economic data and trip generation rates, and simulating the propagation of their variation through a series of common demand models. More specifically, Monte Carlo simulation and sensitivity analysis were used to investigate error propagation over an 818-link network covering a 25-zone area of the Dallas-Fort Worth metro region. The results suggest that uncertainty is likely to compound itself – rather than attenuate – over a series of models. Whilst counteracted to some degree by equilibrium assignment, mis-predictions at early stages (e.g. trip generation) in multi-stage models appeared to amplify across later stages.

4.4 De Jong et al. (2007)

The paper is motivated by a recognition that, for decision-making on infrastructure projects and transport policy measures, it is important to estimate not only the most likely outcome, but also the possible range of future values for traffic volumes and the probabilities attached to these possible outcomes. This is because it might be preferable to invest in a project that is low profit/low risk as opposed to higher profit/higher risk.

De Jong et al. noted that, relative to the plethora of papers on traffic forecasting, there is dearth of papers on quantifying the uncertainty inherent within those forecasts. Collating 21 papers on the topic of uncertainty from the period 1980 to 2007, they summarised for each study:

- The type of uncertainty studied;
- Variables for which uncertainty was studied;
- Methods to quantify uncertainty;
- How uncertainty was expressed;
- Order of magnitude of uncertainty.

For each such study, the authors distinguished between:

a) Input uncertainty: where the future values of the exogenous variables (e.g. the future incomes) are unknown.
b) Model uncertainty: where there is specification error in the model equations (e.g. omitted variables, inappropriate assumptions on functional form and statistical distributions for random components); and error due to using parameter estimates instead of the true values.

De Jong et al. observed that, in order to quantify the degree of input uncertainty, all relevant papers employed some form of repeated model simulation (i.e. sensitivity testing). Typically, statistical distributions were postulated for the input variables and random draws were taken from these distributions. This generated input values which were then used in model runs. The degree of uncertainty was calculated from the variance over all runs for the different input values. Most studies employed univariate distributions for the input variables, thus ignoring correlation between inputs (which would go hand-in-hand with uncertainty propagation).

Turning to model uncertainty, De Jong et al. found there to be a wider range of methods in use. A few studies derived analytical expressions for the variance of the endogenous variable resulting from using parameter estimates for the influence of the exogenous variables. However, for complicated models, these expressions became very cumbersome. It was noted that Jack-knife and Bootstrap methods could be used to elicit proper t-ratios or standard errors for model coefficients in situations with specification error. These (more correct) standard errors of the parameters could then be used in either the analytical calculation of the standard error of the model outcomes or as information on the statistical distributions from which values could be drawn for model simulation runs. Again, the importance of taking account of correlations between parameter estimates was noted.

Following their review of literature, De Jong et al. developed a case study involving the quantification of uncertainty in traffic forecasts from the Dutch national model system LMS. They used existing time series as the key source of information on means, standard deviations and correlations of input variables, and applied these to generate multivariate distributions for the model input variables, thereby accounting for correlation between the input variables. Analytical methods to quantify uncertainty in the tour frequency and mode-destination sub-models of the LMS were considered and the relevant expressions derived, but the computation of these expressions would have required considerable computer time. Instead, the Bootstrap method was used to correct for specification error and Monte Carlo simulation to deal with uncertainty due to estimation.

Both the input variables and the model coefficients of the LMS tour frequency and mode-destination models were varied. Half of the model runs were for the reference situation 2020, and the other half for the situation with a specific road project (namely the extension of the A16 near Rotterdam). De Jong et al. reported uncertainty margins for predictions of a) tours, b) passenger kilometres, and c) vehicle flows, with standard deviations of around 4-16% depending on mode and link. The contribution of input uncertainty (e.g. in future incomes or car ownership levels) to these errors was generally much larger than that of model uncertainty (e.g. coefficients estimated with some error margin).

4.5 Rasouli & Timmermans (2012)

Complementing De Jong et al., Rasouli & Timmermans presented a similar summary of uncertainty inherent within 14 travel demand forecasting studies from the period 2002 to 2012. They grouped these studies into four types of models, namely: traditional four-stage models; discrete choice models; activity-based models (microsimulation); and activity-based models (computational process models). For each study they summarised:

- Type of uncertainty
- Sample fraction and number of model runs
- Model outputs
- How uncertainty was measured and magnitude thereof

Emanating from this review, Rasouli & Timmermans drew the following conclusions:

- Relative to many other topics in the transportation research community, the issue of uncertainty has received only minor attention.
- The effort invested in understanding modelling uncertainty has varied widely by the kind of research. Studies have focussed upon forecasting models of travel demand; with little attention devoted to analytical studies addressing the strength and nature of the relationship between facets of travel demand and their covariates.
- Most studies have been rather ad hoc, in the sense that often only a single source of uncertainty has been examined, and little effort has been invested in systemically varying in a more sophisticated way the factors of interest.

- Some topics have received somewhat more attention. For example, the question of how many simulation runs are required to achieve stable moving averages has been studied by several authors. Several studies have also examined error propagation, especially in the context of four-stage models; these studies have often found that uncertainty increases in the first three steps, before reducing again at the assignment step due to capacity constraints.

Looking to the future research agenda, Rasouli & Timmermans made the following recommendations:

- To date, uncertainty analysis has mainly been concerned with aggregate travel indices such as vehicle miles travelled, O-D matrices and link traffic flows/volumes. By contrast, little attention has been devoted to the various choice facets underlying activity-based models of travel demand and to individual space-time trajectories. Destination and transport mode choice are key components of non-activity-based models as well, but for some reason do not seem to have drawn much attention in assessing uncertainty in travel forecasts.
- Most studies of uncertainty have assumed univariate or multivariate probability distributions, sometimes with covariance terms to represent input uncertainty. This may be a realistic assumption for some error generating processes, but not for others. For instance, most activity-based models of travel demand will use free flow travel times as input, such that the actual probability distribution of travel times will likely embody a positive skew. Moreover, a valuable line of research would be to compare alternative probability functions for their respective impacts on uncertainty.

4.6 Yang et al. (2013)

As the context for their paper, Yang et al. observed that both transportation planning and project evaluation rely upon travel demand forecasting, which is subject to uncertainties stemming from predicted socio-economic inputs, calibrated parameters, and the travel demand model itself. Without considering uncertainty in travel demand modelling, transportation planning and project evaluation may entail unnecessary risk, and any decisions based on such modelling may be inaccurate and misleading.

Yang et al. distinguished between three steps in analysing such uncertainty, namely:

a) Analysing the distributional characteristics of ‘input’ and ‘parameter’ uncertainty in the model, where the former reflects inherent uncertainty in model inputs, whilst the latter reflects uncertainty in parameters which could potentially be reduced by collecting more and/or better quality data.

b) Analysing the manner in which model input/parameter uncertainty is propagated by the model into model ‘output’ uncertainty.

c) Analysing the distributional characteristics of the output uncertainty, for given inputs and parameters.

The objective of Yang et al.’s study was to develop a systematic network equilibrium approach for quantitative uncertainty analysis of a combined travel demand model (CTDM) using the ‘analytical sensitivity-based method’. The CTDM is based on random utility theory, which Yang et al. claimed is ‘behaviourally richer’ than the four-stage model, and can be formulated as an equivalent convex optimisation problem, which lends itself to sensitivity analysis. According to Yang et al., the analytical sensitivity-based method incurs significantly less computational effort then sampling-based methods, and uncertainties stemming from inputs and parameters can be treated separately, such that the individual and collective effects of uncertainty on the outputs can be clearly assessed and quantified.

Illustrating by means of a worked example, Yang et al. found that, at each disaggregate choice step except for the travel choice step, the impact of parameter uncertainty on the output uncertainty was generally more important than input uncertainty. The paper concluded that such insights could allow planners to more effectively allocate limited resources to input data collection and parameter estimation of key variables.

4.7 Wheat & Batley (2015)

Emanating from research funded by DfT, Wheat & Batley (2015) examined the specific uncertainty inherent within appraisal values of travel time savings (VTTS) for non-business travel in WebTAG. Whilst the paper is based upon the WebTAG approach to VTTS pre-2016 (i.e. prior to the recent major update to guidance), the principles remain entirely relevant to present interests. The specific objectives of the paper were:

- to quantify the uncertainty associated with official UK estimates of non-work VTTS; and
- within these estimates, to identify key sources of uncertainty.

In general terms, Wheat & Batley highlighted two key dimensions to the analysis of uncertainty in VTTS estimates.

a) First and foremost, the precision of base year estimates of VTTS, which typically arise from discrete choice modelling of willingness-to-pay data.
b) Second, given the 30-60 year economic life that typifies transport investment schemes, any imprecision in base year estimates of VTTS may be compounded as the scheme proceeds through its economic life.

Within the non-work VTTS formula employed in the UK pre-2016, the latter source of uncertainty was associated with the scaling of base year estimates by a GDP multiplier and associated GDP elasticity, which arose from a further econometric 'meta' model of Stated Preference (SP) and Revealed Preference (RP) evidence on VTTS. According to this relationship, as GDP moves away from the base year level, so the uncertainty in the estimate of the GDP elasticity is amplified within the overall non-work VTTS estimate; this in turn increases the error associated with the VTTS estimate.

Wheat & Batley's focus was the computation of interval estimates for appraisal VTTS for non-business travel. The paper has important conclusions concerning the relative benefits – in terms of uncertainty in appraisal VTTS – of resampling the base year VTTS vs. improving the precision of the income elasticity in the uprating equation. Importantly, it is shown that the interval widths increase dramatically as VTTS is forecast further into the future. This has a significant modelling implication in that it is the uncertainty associated with the process of uprating base estimates of VTTS which results in the large interval width rather than that associated with the base year VTTS estimate. This in turn implies a need to regularly resample the base VTTS, so as to avoid excessive extrapolation of the base VTTS into the future.

4.8 Rezaeestakhruie (2017)

Rezaeestakhruie's recent PhD thesis was motivated by the observation that transportation demand models currently lack a rigorous and analytic treatment to quantify the error propagation from different sources through the models. In common with other papers in the area, the error of traffic forecasts is attributed to two main sources: the model specification error and the input variable measurement error. Rezaeestakhruie observed that, whilst the traditional four-stage transport model is commonly used in practice, the error inherent within the model is not commonly analysed.

Against this background, the first part of the thesis illustrates how the errors of the input variables as well as of the model specification are propagated analytically step-by-step, and how these errors interact to result in inaccurate traffic forecasts. An approach is proposed to quantify separately and collectively the contributions of different sources of error to the overall traffic forecast error. This approach is claimed to be more efficient than existing simulation-based approaches, making it amenable to the analysis of input measurement error and the quality of modelling in large scale networks. The proposed approach derives the variance from calibrated models in each of the four stages, to obtain the variance of the output based on the variance of inputs. The resulting variance formula provides an analytical expression to estimate the forecast errors from the input errors. In addition, the model specification error of each step of the FSTDM is added to the propagated input measurement errors.

This approach was applied to the case study of Brisbane, which involved developing four-stage models for each of eight trip purposes. As an example, a measurement error of 10 percent for the input variables of the Brisbane FSTDM, in combination with specification errors of models calibrated for the Home Based Work – Blue collar (HBWB) trip purpose, was explored. The model specification error produced variances of 1760.77 $(trip/h)^2$, 976.72 $(trip/h)^2$, 0.01082 $(trip/h)^2$ and 0.001327 respectively, for trip production, trip attraction, trip distribution and modal split steps. Subsequently, the variance of output errors for the same steps were 2885.50 $(trip/h)^2$, 7218.70 $(trip/h)^2$, 0.25 $(trip/h)^2$ and 0.18 respectively. The variance of output error in the traffic assignment step was calculated to be 2097.20 $(veh/h)^2$ for all trip purposes, whilst the model specification error of the same step was 1056 $(veh/h)^2$. Given the existing 868 traffic zones, a reduction in the variance of trips per origin-destination (O-D) pair was observed in moving from the first to the third step. At the same time, in the traffic assignment step, considering all trip purposes, the size of the forecast error variance per link increased.

The second part of the thesis deals with the measurement of specification error from a user equilibrium traffic assignment (UETA). More specifically, the propagation of O-D demand measurement errors to the UETA output was investigated using two different methods: the proposed analytical sensitivity-based method and also a simulation-based method. The analytical method employed the results of a sensitivity analysis on the UETA mathematical program, whilst the simulation-based method employed a Monte Carlo Simulation (MCS). The proposed analytical method for error propagation was applied to an empirical example, so as to inform three main questions: the number of samples that ensure a reasonably accurate result for the MCS method; the size of the O-D demand measurement error for which the analytical method remains valid; and the share of the path flow rate variance and covariance from the variance of the O-D demand measurement error.

4.9 Some insights from the literature

Whilst reiterating that this review has (given time constraints) been necessarily selective, the literature surveyed would seem to exhibit a reasonable degree of consensus in terms of the key concepts relevant to the treatment of uncertainty in modelling and appraisal. Moreover, the literature encompasses distinct notions of uncertainty analysis, as follows.

- Uncertainty in appraisal

Pearman considered how appraisal should account for inherent uncertainty in future states of nature. Following Arrow & Lind (1970), the conventional means of dealing with this would be through taking expectations of net benefits and introducing a risk premium, thereby deriving expected NPV. In this context, the discount rate[2] is conventionally 'risk free' – otherwise there would be double-counting with the risk premium. However, this convention is potentially in conflict with current Treasury advice that: *'where the appraisal of a proposal depends materially upon the discounting of effects in the very long term, the received view is that a lower discount rate for the longer term (beyond 30 years) should be used. The main rationale for declining long-term discount rates result is from uncertainty about the future'*. Also pertinent to appraisal, Wheat & Batley (2015) considered the uncertainty inherent within unit valuations of a key source of transport benefit, namely time savings. In this case, the valuations were influenced by both input and modelling error, and the manner in which these sources of error propagated through the VTTS model.

- Uncertainty in modelling and forecasting

The remaining papers considered the uncertainty inherent within traffic and/or benefit forecasts generated by different forms of transport model, again encompassing input and modelling errors, and their propagation. In this context, the literature reveals three interesting findings:

a) There would seem to be consensus that, in aggregate terms at least, error is amplified rather than attenuated as it passes through a modelling system.
b) Notwithstanding the previous point concerning aggregate error, there is evidence that, in traditional four-stage models, the assignment step will tend to attenuate error, whilst the other three steps will tend to amplify error.
c) There would seem to be a lack of consensus on whether input error exceeds modelling error, or vice versa. That is to say, the relative magnitudes of these errors would seem to be context specific.

[2] When conducting appraisals, governments (or private sector investors) employ 'discounting' to take account of how costs, revenues and benefits are distributed over time, for at least the following reasons:

- Resources used to produce a future benefit could be invested elsewhere
- Diminishing marginal utility of income
- Whilst most people care about future populations, they may care less about distantly future populations
- There is a chance that future benefit or cost may not occur

5. IN WHAT WAYS DOES THE DEPARTMENT'S GUIDANCE ALREADY ACCOUNT FOR UNCERTAINTY?

The concept of uncertainty (broadly defined) is represented in various parts of current WebTAG guidance. This section details the principal such references – although there are other more minor references not covered here.

5.1 TAG Unit A1.1: Cost Benefit Analysis

TAG Unit A1.1 deals with 'Cost Benefit Analysis'. This includes a brief section (2.9) entitled 'uncertainty and sensitivity testing' which refers the reader to TAG Unit M4.

Specifically, the text reads: *'TAG Unit M4 – Forecasting and Uncertainty provides guidance on alternative scenarios that should be modelled as sensitivity tests to reflect uncertainty in local factors and national demand growth. The principles described above are equally applicable to alternative scenarios as they are to the core scenario'*.

TAG Unit A1.1 also cross-references TAG Unit A1.2, which *'provides guidance on estimating scheme investment and operating costs, including on applying adjustments for risk and optimism bias'*.

5.2 TAG Unit A1.2: Scheme Costs

TAG Unit A1.2 deals with 'scheme costs', and this includes a reasonably detailed treatment of 'cost risk and uncertainty'. By way of motivation, the guidance notes that:

'Risk in the context of this unit refers to identifiable factors that may impact on scheme costs, leading to over- or under-spends. Such risks should be identified and quantified in a Quantified Risk Assessment (QRA) to produce a risk-adjusted cost estimate. This is required for all transport projects with a base cost greater than £5m in 2010 prices, and is encouraged for smaller schemes'.

The guidance also quotes from the Green Book, in noting that there exists *'a demonstrated systematic, tendency for project appraisers to be overly optimistic'* HMT Green Book (2003, p29). In response to this, TAG Unit A1.2 recommends that *'as well as adjusting for identified, quantified risks, risk-adjusted scheme costs should be adjusted to take account of this optimism bias'*.

5.2.1. Quantified Risk Assessment (QRA)

TAG Unit A1.2 outlines a four-step approach to QRA, as follows:

1. Risk identification: this entails the construction of a comprehensive risk register identifying any risks that are likely to affect the delivery and operation of a scheme.
2. Assessing the impacts of risks to determine possible outcomes: the next step is to assess the cost outcomes of each identified risk.
3. Estimating the likelihood of the outcomes occurring: this involves assigning a probability to each of the possible outcomes.
4. Deriving the probability distribution for the costs of the scheme: having determined each of the possible cost outcomes and their probabilities of occurring, the final step is to derive the complete distribution of cost outcomes.

In essence, QRA accounts for cost risk by analysing the distribution of costs, thereby facilitating derivation of measures of central tendency (e.g. expected cost), as well as the second moment of the distribution.

TAG Unit A1.2 does not impose specific methods for implementing the four steps, and the user is therefore given licence to impose different distributional assumptions, and elicit different measures of the first and second moments. An illustration is however given (as below) which shows a discrete representation of the relevant PDF, as well as the associated CDF showing costs at the mean and 50th and 80th percentiles.

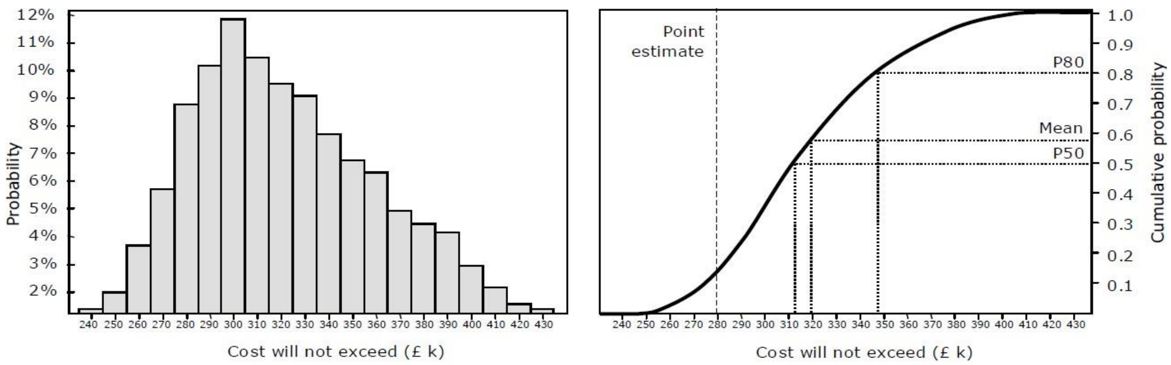

Figure 1 Example Probability Distribution for the costs of a Scheme

(Source: DfT (2017) TAG Unit A1.2)

5.2.2. Optimism bias

TAG Unit A1.2 also outlines a four-step approach to optimism bias, as follows:

5. Determine the nature of the project: this involves categorising the relevant project according to a four-part typology (rail, fixed links, building projects, or IT projects).
6. Identify the stage of scheme development: this is determined in accordance with a three-stage process (where the stages are detailed differently for local/PT, HE and railway schemes).
7. Apply the recommended uplift factors to the risk adjusted transport cost estimate: TAG Unit A1.2 then gives recommended optimism bias uplifts for each category of project/stage of scheme development (although the user is given licence to depart from these recommendations and employ other evidence-based uplifts).
8. Provide sensitivity analysis around the central estimate: the final step requires sensitivity testing around the uplift used.

5.3 TAG Unit A1.3: User and Provider Impacts

This unit includes brief discussion of modelling uncertainty in relation to values of travel time savings, as follows.

'There is a significant degree of uncertainty around the values of travel time savings in the course of work and analysts should undertake sensitivity tests to demonstrate the sensitivity of the appraisal results to the value used'.

'Based on the results of the most recent value of time research, a sensitivity test of +/- 25% around the values of time should be carried out. This represents the average size of the 95% confidence interval around the VTTS across all trip distances. For simplicity and proportionality, this test can be applied as an adjustment to the present value of time saving benefits for business travellers'.

5.4 TAG Unit A2.2: Regeneration Impacts[3]

This unit includes a brief discursive section (4.0) on 'identifying risks and uncertainties in assessment of regeneration impacts'. In particular, two types of uncertainties are identified:

1. Uncertainty about the current situation: *'One source of uncertainty is the quality of information and data available when the assessment is prepared. The data that is used may be an approximation, out of date, inaccurately measured, or be no more than a rough estimate. The assessment of regeneration impacts should comment on the reliability of the data and, where possible, provide ranges, rather than single values, for numeric outputs'*.
2. Uncertainty about real decisions: *'A transport scheme can only provide opportunities for new economic potential that others may then choose to exploit or not. Uncertainties regarding whether these economic opportunities will be exploited, to deliver the regeneration impacts expected, should be identified and examined. Where these impacts are modelled there may be uncertainty about model parameters'*.

No specific guidance on dealing with these uncertainties is given.

5.5 TAG Unit A3: Environmental Impact Appraisal

This unit outlines a five-step approach to environmental impact assessment, as follows:

- Step 1: Scoping and identification of study area
- Step 2: Identifying key environmental resources and describing their features
- Step 3: Appraise environmental capital
- Step 4: Appraise the proposal's impact
- Step 5: Determine the overall assessment score

TAG Unit A3 notes that: *'Steps 2 to 4 of the appraisal may have a risk component, where the exact impacts of the scheme are unknown because of uncertainties in exposure and effect. Where uncertainties of this sort are identified, they should be made explicit in the appraisal process. It is recommended that the precautionary principle be employed. Even at larger scales where there is likely to be greater uncertainty regarding the potential impacts, there remains the opportunity to incorporate mitigation measures when the schemes are considered in more detail. In these cases it will be necessary to determine whether the potential risks identified justify invoking the precautionary principle, or whether it will be sufficient to flag them up as issues for more detailed consideration at a later stage'.*

[3] Note that TAG Unit A2.2 is soon to be withdrawn and assimilated into the wider economic impacts guidance.

5.6 TAG Unit A4.1: Social Impact Appraisal

This TAG unit encompasses a range of areas. One such area is health and safety risks, but these are considered to be outside the scope of the current stocktake. Another area, which is more relevant to the stocktake, is option and non-use values. The guidance notes that: *'Option and non-use values should be assessed if the scheme being appraised includes measures that will substantially change the availability of transport services within the study area (e.g. the opening or closure of a rail service, or the introduction or withdrawal of buses serving a particular rural area)'*. The guidance proceeds to highlight some important features of option values, namely:

- '*They are associated with uncertainty about use of the transport facility;*
- *They may exist even if the option of using the transport service is never taken up;*
- *They are related to the individual's attitude to uncertainty'.*

No specific guidance is given on how option and non-use values should be quantified.

5.7 TAG Unit A5.3 Rail Appraisal

Whilst guidance in TAG Unit A1.1 on 'Cost Benefit Analysis' provides the facility for scheme promoters to extrapolate the impacts of demand growth beyond the final year forecast, many road scheme promoters in practice cap demand growth 15-25 years after the scheme opening year. Furthermore, until very recently, TAG Unit A5.3 on 'Rail Appraisal' recommended the imposition of a 'demand cap' beyond 20 years from the appraisal year. The demand cap concept offers a pragmatic means of dealing with uncertainty over the longer-term. However, DfT has become increasingly uncomfortable with the implication that, where population growth continues, imposition of the demand cap will produce forecasts that assume decline in per capita rail travel – when this is not supported by actual market trends.

Against this background, Section 2.3 of A5.3 deals with 'demand and revenue forecasting', and begins by outlining the 'standard' approach to forecasting based on exogenous elasticities recommended in TAG Unit M4. However, in a recent update to this unit, a distinction is made between two phases of the forecasting horizon, as follows:

- *'…models using the standard approach should be used to provide a forecast twenty years from the appraisal year (e.g. an appraisal carried out in 2017/18 would be on the basis of a 2037/38 forecast)'*.
- *'Beyond this point, and for the remainder of the appraisal period, the magnitude of impacts associated with the level of demand may be extrapolated using a simple forecast model, in line with population growth'*.

In applying this approach, the guidance qualifies: *'This assumes that the magnitude of impacts per capita remain fixed at the levels observed in the final forecast year, but that journeys (and therefore total impacts) continue to increase as population increases. Before applying such extrapolation, analysts are expected to consider whether capacity limitations are likely to constrain the impacts not just within the standard twenty year modelling horizon but also over the remainder of the appraisal period, for example because of over-crowding in the do something'*.

Where population-based extrapolation is used, the guidance issues a requirement for sensitivity testing.

5.8 TAG Unit M1.1: Principles of Modelling and Forecasting

This TAG unit issues separate guidance on the mitigation of modelling and forecasting risks, as follows.

5.8.1. Mitigating modelling risks

In Section 3 of this TAG unit, it is noted that:

'There is a risk that model may not be realistic or sensible due to the error around the model parameters used, or limitations in the extent to which the model can represent human behaviour. Therefore, before using any mathematical model, it is essential to check that it produces credible outputs consistent with observed behaviour. This is usually done by running the model for the base year (either the current year or a recent year), and:

- *comparing its outputs with independent data (validation);*
- *checking that its response to changes in inputs is realistic, based on results from independent evidence (realism testing); and*
- *checking that the model responds appropriately to all its main inputs (sensitivity testing)'*.

The guidance goes on to discuss several specific modelling risks, and possible mitigations, as follows:

- Mitigating the risk of errors in model inputs: *'Inputs to transport models should be transparent and straightforward to audit'*.
- Using models in accordance with their design and underlying theory: *'Any model is a simplification of reality. Although it is often sensible to use existing models to save unnecessary costs, a model designed for one purpose may not be suitable for a different situation…There may also be circumstances where it is technically possible to obtain results for appraisal whilst using a model in a way inconsistent with its underlying principles. This can lead to misleading appraisal results'*.
- Achieving an appropriate representation of human behaviour: *'The quality of the model response should be tested by running the model for a designated base year (usually the current year or a recent historic year) and carrying out the following tests:*
  - *validation: comparing model outputs with observed data;*
  - *realism testing: rerunning the model with some standard changes to inputs, such as fuel prices, public transport fares and car journey time, to check that the model responses (elasticities) are realistic; and*
  - *sensitivity testing: rerunning the model with changes to model parameters, to check the model results are robust to changes in these parameters (or otherwise indicate areas of risk if the model inputs are changed)'.*

5.8.2. Mitigating forecasting risks

Section 5 of this TAG unit deals with forecasting risk, distinguishing between sources of bias and uncertainty.

a) Sources of bias

The guidance notes: *'Transport schemes often have both positive and negative impacts, both of which are usually augmented as demand for the transport schemes increase. It is therefore not possible to create a universal "worst-case" scenario that takes into account all risks. Instead, the primary basis of evidence should be the core scenario, which should be developed using unbiased and realistic assumptions'.*

Furthermore: *'The core scenario should also be unbiased with regard to local sources of uncertainty. Such sources of uncertainty therefore need to be identified, through construction of an uncertainty log, as part of the definition of the core scenario'*.

b) Sources of uncertainty

The guidance notes: *'Forecasts are, by nature, uncertain. Even though the assumptions in the core scenario should be unbiased, there is no guarantee that outturn real-world result will be the same as the forecasts in the core scenario'*.

Furthermore: *'Decision-makers need to understand these risks, so it is important for analysts to communicate them well and quantify them if proportionate. Given the complexity of interactions between demand and supply in transport systems, the best way to quantify these risks is to define alternative scenarios using different assumptions to the core scenario, and then re-run the model using these different assumptions'*.

5.9 TAG Unit M4: Forecasting and Uncertainty

Finally, across all of WebTAG, Unit M4 gives perhaps the most focussed account of how uncertainty should be treated in modelling and appraisal. Section 2 of this unit deals with *'uncertainty and the uncertainty log'*, noting that there are two sources of forecast error:

- *'uncertainty in the inputs (such as size of new housing development) and;*
- *error in the model parameters and specification (how these inputs propagate through the model)'*.

The practitioner is advised to: *'summarise all known assumptions and uncertainties in the modelling and forecasting approach in an uncertainty log. The uncertainty log will also be the basis for developing a set of alternative scenarios. The alternative scenario is used to understand the possible impact of an error in assumptions on the model forecasts'*.

The guidance proceeds to give additional guidance on the uncertainty log, as follows.

'The purpose of the uncertainty log is to record the central forecasting assumptions that underpin the core scenario and record the degree of uncertainty around these central assumptions', in relation to the following:

- Model parameter error: this is determined from sensitivity tests, as described in TAG Unit M2;
- National uncertainty in travel demand: this is due to uncertainty in demographic projections and travellers' behaviours and tastes[4];
- National uncertainty in travel cost: this is associated with uncertainty in fuel prices or government policy;
- Local uncertainty (within the vicinity of the scheme) in travel demand: the most common cause here is uncertainty around whether proposed developments (for example housing, employment, schools, or retail) are built;
- Local uncertainty (within the vicinity of the scheme) in travel supply/cost: this could be caused by uncertainty around whether other transport construction projects will materialise.

5.10 Summary of coverage of risk and uncertainty in WebTAG

As summarised in Figure 5.1, existing WebTAG guidance considers various dimensions of uncertainty – but arguably in a rather ad hoc and uncoordinated manner across the modelling and appraisal process.

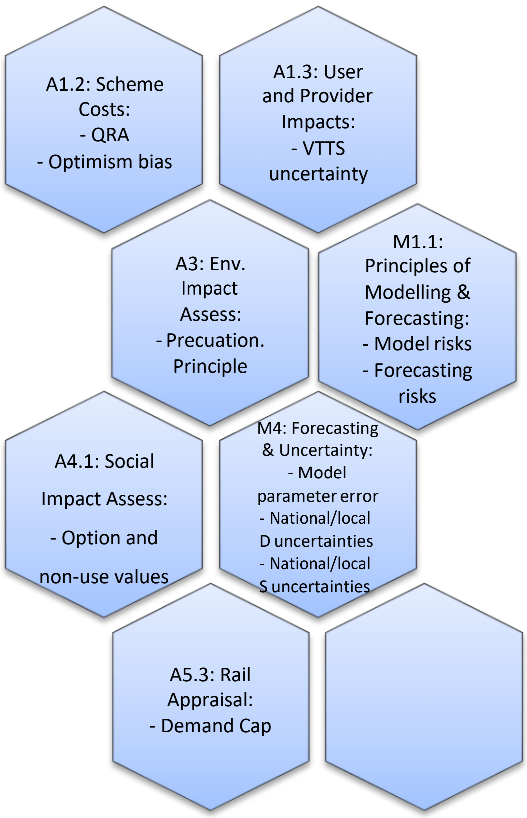

Figure 1 WebTAG units which consider uncertainty in some form and at some substantive level of detail

In Figure 5.2, an attempt is made to organise these various discussions of uncertainty, with reference

[4] In the core scenario, it is assumed that the impact of changes in demographic data will be based on the NTEM dataset, whilst growth in most other parameters will be based on the values given in the TAG Data Book.

to:

- the key stages of the modelling and appraisal process, and
- the key sources of uncertainty identified earlier in this note

To these ends, it might be seen that:

- **MODELLING** (i.e. estimation of a model under DN conditions): The specialist modelling units M1.1 and M4 outline different sources of modelling risk and discuss parameter uncertainty (which could be associated with estimation and/or specification error) in greater depth.
- **FORECASTING** (i.e. application of an estimated model to DM/DS conditions): This would seem to be the principal focus of WebTAG's treatment of uncertainty, with the specialist modelling units M1.1 and M4 outlining different sources of forecasting risk and devoting more detailed attention to national and local sources of uncertainty on both demand and supply sides. Unit A5.3 describes the 'demand cap' method of dealing with demand-side uncertainty, whilst A1.2 describes methods for dealing with uncertainty in scheme costs on the supply-side. Finally, A3 considers uncertainties associated with environmental impacts.
- **APPRAISAL** (i.e. the social benefits/costs of the DS vs. DM or DN): The appraisal stage of the process combines the modelling of the DN and forecasts of the DM/DS with assumptions (ideally evidence-based) concerning user benefits/costs – such that the aggregate benefits and costs of a given scheme can be estimated. Within this context, Unit A1.3 offers some discussion of uncertainty in values of travel time savings (VTTS) – which is arguably the single most important source of user benefit in transport appraisal. Finally, A4.1 gives guidance on option and non-use values.

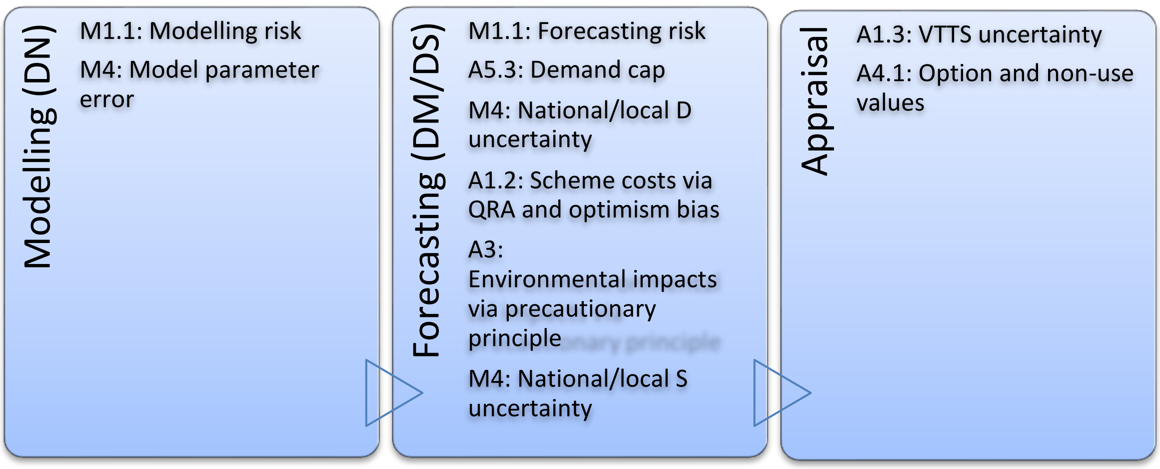

Figure 5.2: How WebTAG units which consider uncertainty inform different stages of the modelling and appraisal process

5.10 TASM's ongoing development of future planning scenarios

Whilst not as yet implemented in WebTAG, TASM has in recent years devoted significant effort to the development of future planning scenarios. In particular, TASM moved to a scenarios-based approach for RTF15, noting that this move was *'well received by external stakeholders and the wider public as it better represented underlying uncertainty in travel demand and enabled Ministers to make more informed decisions'* (TASM internal note). Subsequently, TASM has sought to extend that process by considering a wider range of underlying sources of uncertainty, with the objective of developing a more definitive set of scenarios which could potentially be adopted across the Department.

5.11 Core scenario

At the time of writing, TASM's 'core' scenario underpinning RIS2 (DfT, 2017) is based upon the 'best available evidence', and to this end employs the same assumptions as NTEM 7.2 (core GDP forecasts; ONS forecasts for population; current trends in age and gender driving patterns; constant trip rates (from 2016); current trends in urbanisation shifts and geographic economic structure, meaning continued urbanisation and continued growth of London at faster rate relative to rest of country). Furthermore, it include BEIS forecasts for fuel, current trends in freight mix (high LGV growth, relatively flat HGV growth) and WebTAG recommendations on VTTS.

5.12 Sensitivities around the core scenario

In tandem with the core scenario, TASM has modelled sensitivity tests around the core which reflect uncertainty about future GDP and fuel price levels.

Beyond these sensitivities, TASM has undertaken work to develop additional scenarios reflecting other key sources of uncertainty, as follows:

- Location: this refers to uncertainty about where population and economic growth will occur in the future, in particular the rate of urbanisation and whether economic growth will be concentrated in London or more distributed across the country.
- People's propensity to travel: this refers to uncertainty about what the driving patterns of different demographic groups could look like in the future.
- Technology: this refers to uncertainty about ongoing technological changes (e.g. Mobility as a Service (MaaS), Connected and Autonomous Vehicles (CAVs) and Ultra Low Emission Vehicles (ULEVs)), and their implications for transport services and behaviours.

Given the significance of some of these changes, these might be seen as structural breaks rather than trend effects – making it difficult to include evidence-based assumptions within forecasting models.

6. PRESENTATION OF UNCERTAINTY

Before responding to the 'stocktake' questions posed by TASM, a final substantive topic considered in this note is the presentation of uncertainty (especially to policymakers). In this regard, the 2015 TASM note to JADP on modelling and appraisal uncertainty commented:

'Currently it is common practice to present the main risks to the project to the decision-maker. This includes:

- *transparency regarding the main assumptions;*
- *the quality of the analysis and the potential analytical assurance risk that this presents;*
- *a range around the core BCR and a description of the likelihood of variance;*
- *what-if scenarios and key factors that may push the BCR into a lower category;*
- *the most relevant sensitivity tests and analysis around the key uncertainties.*

Of course, the most appropriate sensitivity tests and presentation of uncertainty will vary from scheme to scheme. However, we have been challenged to be more proactive in considering better ways to provide a richer analytical picture, and to highlight key risks, when presenting business cases to investment boards'.

The TASM note highlighted three presentational devices which have been used in recent DfT policy documents, namely:

a) The profiling of future scheme benefits that would arise under a 'demand cap' (Figure 6.1), as explained earlier in Section 5.7.
b) The distributional properties of scheme BCR as calculated through 'Monte Carlo' simulation (Figure 6.2). Monte Carlo methods are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results – they are ideally suited to the problem of enumerating uncertainty in modelling outcomes.
c) The range of national traffic forecasts under 'scenario analysis' (Figure 6.3), as detailed earlier in Section 5.11.

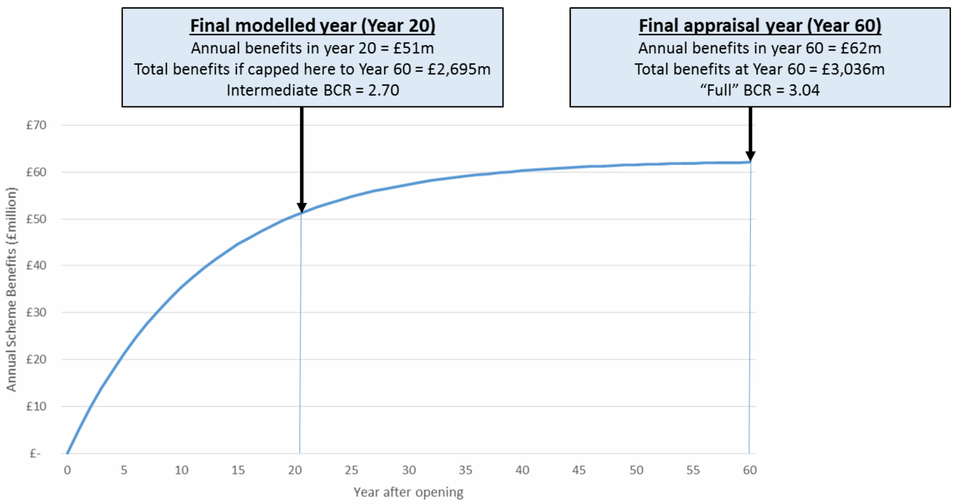

Figure 6.1: Future profile of scheme benefits via a demand cap (Source: DfT (2015) JADP paper on long-term forecasting and uncertainty)

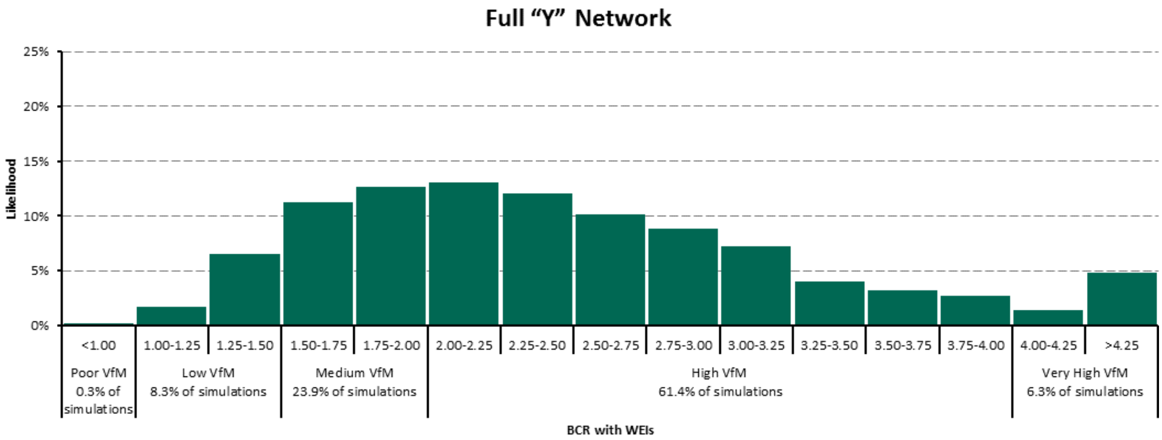

Figure 6.2: Range of potential BCRs via Monte Carlo simulation (Source: DfT (2015) HS2 Phase 2 Economic Case)

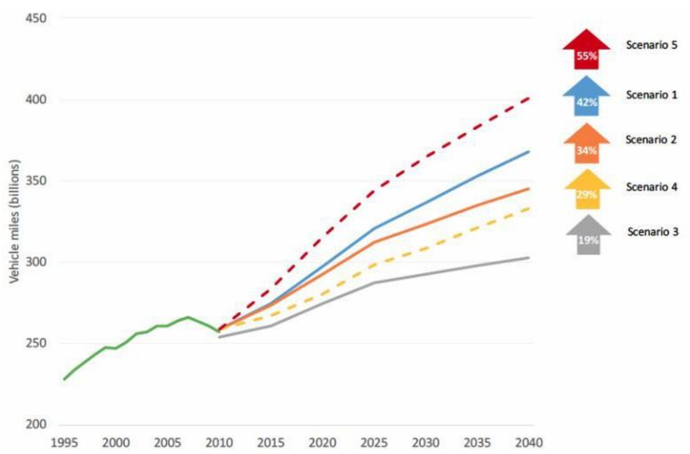

Figure 6.3: Range of national traffic forecasts via scenarios approach (Source: DfT (2015) RTF)

In a similar spirit, one might also note the earlier figures in Section 5.2.1 above, giving the PDF and CDF associated with a QRA of scheme costs.

Whilst the above devices offer various means of representing modelling and appraisal uncertainty, they are neither exhaustive of the presentational possibilities, nor mutually exclusive. For example, Figure 6.4 combines the notions of both range and scenario to develop the so called 'flying bars' presentation.

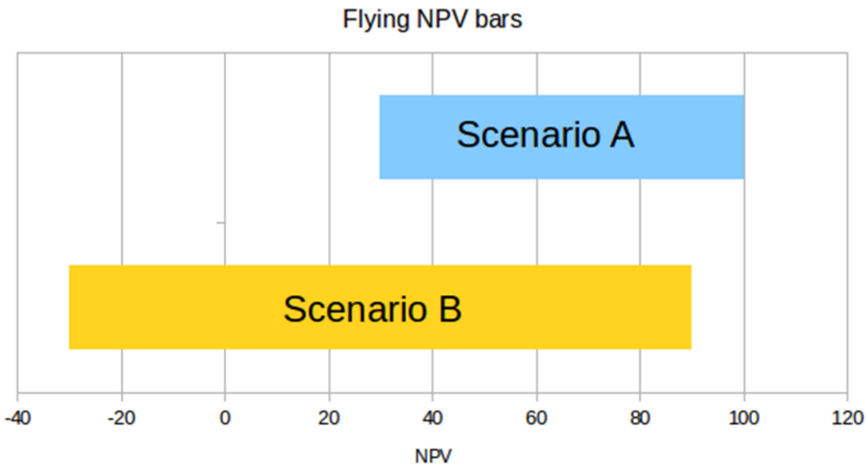

Figure 6.4: Example of 'flying NPV bars' (Source: ITF (2017) Strategic Infrastructure Planning – International Best Practice)

Another approach is the 'risk analysis' presentation shown in Figure 6.5. Risk analysis is a method applied to individual projects in order to determine the relative importance of different variables as determinants of project returns. In particular, the method estimates the likelihood that project returns (e.g. expected internal rate of return, EIRR) are unacceptable. To this end, Figure 6.5 generates a CDF of project returns, and identifies the probability that the EIRR will fall below and above specified thresholds of acceptability.

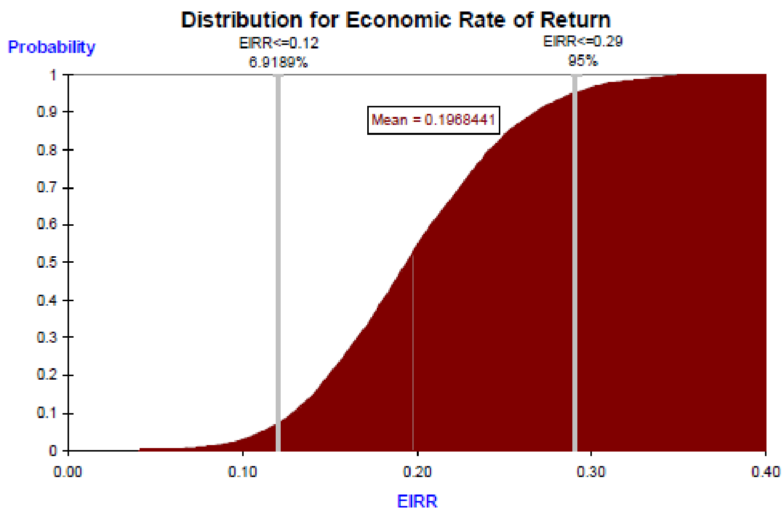

Figure 6.5: Distribution of economic rate of return via risk analysis approach (Source: ADB (2011))

7. RESEARCH PRIORITIES AND ACTIONS TO ADDRESS THESE PRIORITIES

The preceding sections of this note have set out the background to the treatment of uncertainty in transport modelling and appraisal. Equipped with this background, this (the final) section responds to several specific requirements detailed by the Department in the scope for this stocktake – especially around the identification of research priorities, and the actions that might be taken to address these priorities.

7.1 Existing coverage of uncertainty within WebTAG

a) *Comment on where there may be gaps in the existing toolkit and guidance (noting that a substantial amount of work on the use of scenarios is already underway)*

With reference to Section 5.10 of this note, it is appropriate to qualify the reference to the 'existing toolkit and guidance'. WebTAG encompasses guidance on various aspects of uncertainty (broadly defined) within the modelling and appraisal process, and the quality of some aspects of this guidance (e.g. on scheme costs) is very respectable. Furthermore, the ongoing TASM work on scenario analysis might be considered best (even leading) practice for national transport ministries. However, in totality, TAG guidance on uncertainty gives the impression of having evolved in a rather ad hoc fashion over time, with the following implications:

- Most fundamentally, there is a lack of consistency/agreement on concepts and terminology – TAG employs the terms 'risk', 'uncertainty', 'error' and 'robustness' (and possibly others) at various junctures, without always being clear as to precisely what concepts are being referred to.
- Different parts of TAG devote different levels of focus to 'uncertainty' (broadly defined), such that the final product gives a fragmented picture at an inconsistent level of detail. This is compounded by the manner in which the guidance is organised – with some generic material located in the specialist modelling unit M4, and other material dispersed around a number of other units.
- Whilst this note has introduced various typologies for characterising uncertainty, there are perhaps two broad areas which should be most pertinent to the Department's interests, namely:
    1) Input uncertainty (encompassing any errors in data, assumptions concerning future planning variables and any assumed parameters)
    2) Model error (encompassing any error in estimated parameters and any specification error)
- Moreover, given the somewhat fragmented and inconsistent treatment proffered by the guidance, it is difficult to discern a clear sense of what exactly constitutes the 'toolkit' for dealing with uncertainty in modelling and appraisal.

b) *Comment on the feasibility of specifying standardised ranges for key demand drivers for use in sensitivity testing (such as GDP, population, employment, energy costs etc)*

Note that this relates primarily to the item 1) above. It would seem eminently feasible to recommend standard ranges for key demand drivers, and these are readily available in some cases (e.g. see variants of future population projections in ONS (2016)).

c) *Comment on the feasibility/desirability of providing ranges for demand model parameters (such as elasticity / choice model coefficient estimates); and*

Note that this relates primarily to the item 2) above. When estimating model parameters, standard parametric techniques typically generate not only a mean estimate but also a standard error. Depending on the precise specification of the model, it may be straightforward to derive a distribution around the mean estimate, but in some cases this process could be complex (and possibly intractable). That said, it would not seem unreasonable to challenge the modeller to report both first and second moments of a given estimate (and indeed this might influence trade-offs between different model specifications at the model development stage). Whilst there should be an onus on the modeller to investigate ranges for model parameters, it would seem rather more difficult to make a priori judgement on the acceptability of any given range – since this will largely be context specific.

d) Comment on the feasibility/desirability of refreshing evidence or capturing uncertainty around evidence used in the calibration of models (such as fuel price elasticities, observed traffic etc)

Input uncertainty could arise from errors in data (such as observed traffic), future planning variables (such as GDP) and/or assumed parameters (such as fuel price elasticities). It would be both feasible and desirable to update model calibrations – in the light of updated and/or better evidence – with the objective of mitigating these three sources of error.

Whilst it is difficult to generalise, it should be borne in mind that actions to reduce these three sources of error would likely incur different resource costs and could reap different dividends in terms of moderating uncertainty. An interesting case in point is the Wheat & Batley (2015) paper, in the context of uncertainty in VTTS forecasts. That is to say, Wheat & Batley considered the benefits of both updating the GDP elasticity (which might be seen as revising a parameter assumption) and re-sampling behavioural estimates of VTTS (which might be seen as collecting new data).

More generally, for key models within the Department's modelling and appraisal suite, there would be considerable value in: i) identifying key sources of uncertainty, and ii) seeking understanding of how these sources propagate through the relevant model. Equipped with these insights, the Department could make informed decisions on the cost vs. analytical benefit of different actions to mitigate input uncertainty.

e) Comment on desirability of making references in the guidance to methods other than standard sensitivity test or scenarios (such as Monte Carlo, switching values or any other techniques)

Given the observations above regarding the coherence of the existing 'toolkit', it is arguable whether the exploration of alternative methods should be a focus of attention – at least not in the short term. Rather the focus should perhaps be on:

- Consolidating existing guidance, with the objective of eliciting a coherent 'toolkit'.
- Enhancing the level of instruction within the 'toolkit'.

On the latter, existing TAG guidance on uncertainty (e.g. M4) is written at a fairly general level, and practitioners are seemingly given considerable licence to deploy various analytical methods in a manner of their choosing. Whilst there is always a balance to be struck on the level of prescription within WebTAG (mindful that additional prescription can increase modelling cost and reduce modelling flexibility), strengthening of guidance on uncertainty would promote greater consistency of approach – and greater rigour.

As regards specific methods of uncertainty analysis, a sensible level of ambition in the short term would be to improve guidance around specific 'standard methods' (including the likes of sensitivity analysis, Monte Carlo analysis, switching values, and scenario analysis), before exploring the potential for alternative methods in the longer term.

f) *Comment on uncertainties in appraisal that might not be captured by any of the above or require a separate consideration; and any innovation/improvements to the presentation of uncertainty that the Department could consider.*

Whilst the existing 'toolkit' on uncertainty could be considerably sharpened, it would seem fairly broad in its consideration of different sources of uncertainty. Against this background, it is arguable whether effort in the short term should be devoted to identifying additional sources of uncertainty – rather the priority should be to improve the clarity of guidance on the sources already identified.

Moreover, in developing TAG guidance in this area, a practical aspiration should be to focus attention (and resource) on key sources of uncertainty – which in due would call for a 'convergent' rather than 'divergent' approach to the research programme.

Turning to presentational considerations, it would seem that the Department's thinking in this area is reasonably well developed, and immediate considerations should be focussed upon which means of presentation are most useful in communicating uncertainty to relevant officials and stakeholders.

7.2 Priorities for future research to strengthen WebTAG's coverage of uncertainty

g) *Suggest a priority list based on the scale and importance of the gap in the currently available toolkit and methods*

1. The priority should be to review – and where necessary refresh – existing WebTAG assets, with the objective of delivering a coherent 'toolkit' to deal with uncertainty in transport modelling and appraisal. There is a presentational dilemma as to whether this topic should be devoted a TAG unit in itself (i.e. in the form of an enhanced M4) or whether it should be devolved to other TAG units – at present, it would seem that WebTAG has adopted aspects of both approaches.
2. The purpose of the 'toolkit' should be to equip practitioners with sufficient guidance to enable them to investigate and report key sources of uncertainty in modelling and appraisal in a systematic and consistent fashion. To these ends, the guidance could usefully encompass the following areas:
    - i) Identify sources of input uncertainty (encompassing any errors in data, assumptions concerning future planning variables and any assumed parameters) and model error (encompassing any error in estimated parameters and any specification error) on both demand and supply sides.
    - ii) Where appropriate, outline acceptable ranges for assumed parameters. Otherwise, issue guidance on reporting expectations with regards to observed variances in data inputs and estimated parameters.
    - iii) Outline methods for quantifying key sources of uncertainty both individually and collectively, and especially the propagation of uncertainty through model and appraisal systems. Depending upon the practical context, these methods could

encompass the likes of sensitivity analysis, switching values, scenario analysis, Monte Carlo simulation, jack-knife/bootstrapping tests, and/or risk analysis. The practical context may be dictated not only by the complexity of the modelling/appraisal task at hand, but also by the degree of rigour required in the uncertainty analysis, and the analytical resources available – for example, sensitivity analysis would offer a relatively quick and simple option, whilst Monte Carlo simulation would be more involved but also more robust.

iv) Discuss approaches to mitigating these uncertainties.

v) Issue guidance on reporting expectations with regards to overall uncertainty in modelling and appraisal forecasts (as per Section 6 of this note).

As and when such guidance is developed and released, a longer term priority (informed by item iii) above) should be to form an overview on the key sources of uncertainty – especially if these are in areas where the Department could take action to mitigate (e.g. commission new research fuel cost elasticities, commission new data collection on VTTS, etc.).

h) Indicate which of the gaps can be closed with relatively little effort (low hanging-fruit); and

The above is an eminently feasible scope for updated guidance, although iii) and iv) might require a certain amount of contemplation.

i) Indicate significant inter-dependencies between different items in the list.

Items i) to v) above essentially constitute a 'linear' set of tasks. There is a degree of inter-dependency between iii) and v), in the sense that the choice of methods for quantifying sources of uncertainty may dictate the methods available for presenting uncertainty. For example, Monte Carlo analysis would yield insight into the full distribution of appraisal outcomes, whereas sensitivity analysis would effectively reveal only discrete points along the distribution.

7.3 Format of the future research programme

j) Suggest which areas can potentially be addressed based on existing evidence or with limited additional research effort (likely with in-house resources)

k) Suggest what recommendations can be undertaken as part of an uncertainty stocktake (potentially with external research, but as part of one compact research project);

l) Suggest which areas may need to be covered by standalone or more in-depth research. Such instances are expected to be driven by the level of technical challenge as well as the likely complexity associated with formulating practical recommendations based on any such work.

Taking these three questions together, a future work programme might be comprised of the following elements:

Work package 1: Draft guidance on the uncertainty 'toolkit'

WP1 would draft standalone guidance note on the 'toolkit' for dealing with uncertainty in transport modelling and forecasting – in the manner of point 7.2 above. This could in due course form the basis of a dedicated TAG unit – or alternatively it could provide a framework for representing uncertainty in a coherent and consistent fashion on a 'devolved' basis across relevant TAG units. This guidance should be clear concerning concepts and terminology – and could usefully distinguish between 1) modelling error and 2) uncertainty regarding the future. The guidance note should cover the entirety of the modelling and appraisal process. In preparing this note, it may be useful to consider relevant practices employed by other national transport ministries, and the extent to which they might add value to UK practices.

The optimal team for this WP would probably be some combination of practitioner (bringing insight on practical context) and academic (bringing insight on methods), but would need active contribution from the Department.

Work package 2: Case study analysis of uncertainty mitigation

The objectives of this WP would be three-fold. First, to test the applicability of the 'toolkit' from WP1. Second, to exploit the 'toolkit' with the objectives of identifying the principle sources of uncertainty and understanding their propagation through the modelling and appraisal system. Third, to mitigate those sources of uncertainty (e.g. through updating data, models or forecasting scenarios).

WP2 should involve analytical and policy 'test beds' drawn from a range of different spatial and modal contexts, for example:

- Local four-stage transport model
- Strategic transport model (e.g. VDM)
- National forecasting model (e.g. NTM)

This task could be performed by a consultant with access to such models, but would also benefit from some academic input. Again, active contribution from the Department would be important – especially where DfT is the custodian of any such model.

Work package 3: Update guidance on the uncertainty 'toolkit' and issue research recommendations

In the light of outcomes from WP2, WP3 would update the draft guidance and issue recommendations on where DfT could best invest research resource (e.g. on data collection vs. model recalibration) to mitigate key areas of uncertainty in modelling and appraisal.

Work package 4: Undertake detailed work to mitigate key sources of uncertainty

WP4 is essentially a contingency. Depending upon the recommendations from WP3, there may be a need for additional commissions to update specific areas of the Department's evidence base. It is difficult to predict whether/what commissions might be needed in this regard. However, in committing to this programme of works to strengthen TAG guidance on uncertainty in modelling and appraisal, the Department should be open to the possibility that the programme will expose particular areas of weakness in its evidence base

Commissioning

In terms of commissioning, WP1-3 could potentially be collated as a single integrated commission. Alternatively, WP1 could be commissioned as a standalone activity in the first instance, and depending on the outcomes, the Department could review its plans for WP2 and WP3. Naturally, the second configuration would give the Department additional procurement flexibility, but at the risk of losing continuity and momentum.

REFERENCES

Asian Development Bank (ADB) (2011) Introductory Course on Economic Analysis of Investment Projects, Session 3.3: Sensitivity and Risk Analysis. Available at: https://www.adb.org/sites/default/files/page/149401/sensitivity-risk-analysis-feb2011.pdf

Advisory Committee on Trunk Road Assessment (1977) (Chairman – Sir George Leitch). Report. HMSO, London.

Arrow, K.J. & Lind, R.C. (1970) 'Uncertainty and the evaluation of public investment decisions'. American Economic Reviews, 60, pp.364-78.

Ashley, D.J. (1980) 'Uncertainty in the context of highway appraisal'. Transportation, 9, pp249-267.

De Jong, G., Daly, A., Pieters, M., Miller, S., Plasmeijer, R. & Hofman, F. (2007) 'Uncertainty in traffic forecasts: literature review and new results for the Netherlands'. Paper presented at ETC 2007.

Department for Transport (2014) Strength in Numbers: the DfT Analytical Assurance Framework. Available at: https://www.gov.uk/government/publications/dft-analytical-assurance-framework-strength-in-numbers

Department for Transport (2015) 'Long-Term Forecasting and Uncertainty'. Joint Analysis Development Panel Discussion Paper.

Department for Transport (2015) Road Traffic Forecasts 2015. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/411471/road-traffic-forecasts-2015.pdf

Department for Transport (2017) High Speed Two Phase Two Economic Case. Available at:

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/634196/high-speed-two-phase-two-economic-case.pdf

Department for Transport (2017) TAG UNIT A1.1: Cost-Benefit Analysis. Available at: https://www.gov.uk/government/publications/webtag-tag-unit-a1-1-cost-benefit-analysis-december-2017

Department for Transport (2017) TAG UNIT A1.2: Scheme Costs. Available at: https://www.gov.uk/government/publications/webtag-tag-unit-a1-2-scheme-costs-july-2017

Department for Transport (2017) TAG UNIT A1.3: User and provide Impacts. Available at: https://www.gov.uk/government/publications/webtag-tag-unit-a1-3-user-and-provider-impacts-march-2017

Department for Transport (2014) TAG UNIT A2.2: Regeneration Impacts. Available at: https://www.gov.uk/government/publications/webtag-tag-unit-a2-2-regeneration-impacts

Department for Transport (2015) TAG UNIT A3: Environmental Impact Appraisal. Available at: https://www.gov.uk/government/publications/webtag-tag-unit-a3-environmental-impact-appraisal-december-2015

Department for Transport (2017) TAG UNIT A4.1: Social Impact Appraisal. Available at: https://www.gov.uk/government/publications/webtag-tag-unit-a4-1-social-impact-appraisal-december-2017

Department for Transport (2017) TAG UNIT A5.3: Rail Appraisal. Available at: https://www.gov.uk/government/publications/webtag-tag-unit-a5-3-rail-appraisal-december-2017

Department for Transport (2017) TAG UNIT M1: Principles of Modelling and Forecasting. Available at: https://www.gov.uk/government/publications/webtag-tag-unit-m1-1-principles-of-modelling-and-forecasting

Department for Transport (2017) TAG UNIT M4: Forecasting and Uncertainty. Available at: https://www.gov.uk/government/publications/webtag-tag-unit-m4-forecasting-and-uncertainty-july-2017

Department for Transport (2017) 'Road demand scenarios for the second roads investment strategy'. TASM Internal Note.

HM Treasury (2011) The Green Book: Appraisal and Evaluation in Central Government. TSO, London. Available at: https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-governent

International Transport Forum (2017) Strategic Infrastructure Planning – International Best Practice. Available at: https://www.nic.org.uk/wp-content/uploads/Strategic-Infrastructure-FINAL-for-web_v2.pdf

Keynes, J.M. (1921) A Treatise on Probability. Macmillan, London.

Keynes, J.M. (1936) The General Theory of Employment, Interest and Money. Macmillan, London.

Knight, F.H. (1921) Risk, Uncertainty and Profit. Houghton Mifflin, Boston.

National Audit Office (NAO) (2013) High Speed 2: A Review of Early Programme Preparation. Report by the Comptroller and Auditor General. HC 124, Session 2013-14. Available at: https://www.nao.org.uk/wp-content/uploads/2013/07/Full-Report.pdf

National Infrastructure Commission (2017) 'Congestion, capacity, carbon: priorities for national infrastructure'. Consultation on a National Infrastructure Assessment. Available at: https://www.nic.org.uk/wp-content/uploads/Congestion-Capacity-Carbon_-Priorities-for-national-infrastructure.pdf

Norvig, P. & Thrun, S. (2018) 'Introduction to Artificial Intelligence'. Udacity. Available at: https://www.udacity.com/wiki/cs271/unit1-notes

Office of National Statistics (ONS) (2016) National Population Projections: 2016-based statistical bulletin. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/compendium/nationalpopulationprojections/2016basedprojections/variants

Pearman, A.D. (1976) Transport investment appraisal in the presence of uncertainty. Transportation Research, 10 (5), pp.331-338.

Rasouli, S. & Timmermans,H. (2012) 'Uncertainty in travel demand forecasting models: literature review and research agenda'. Transportation Letters, 4 pp55-73.

Rezaeestakhruie, H. (2017) 'Analytical error propagation in four-step transportation demand models'. Thesis submitted for the degree of Doctor of Philosophy at The University of Queensland in 2017, School of Civil Engineering.

Wheat, P. & Batley, R. (2015) 'Quantifying and decomposing the uncertainty in appraisal values of travel time savings'. Transport Policy, 44, pp134-142.

Yang, C., Chen, A., Xu, X. & Wong, S.C. (2013) 'Sensitivity-based uncertainty analysis of a combined travel demand model' The 20th International Symposium on Transportation and Traffic Theory (ISTTT 2013). Procedia - Social and Behavioral Sciences, 80, pp395-415.

Zhao, Y. & Kockelman, K.M. (2002) 'The propagation of uncertainty through travel demand models: an exploratory analysis'. Presented at the 80th Annual Meeting of the Transportation Research Board. Annals of Regional Science, 36 (1), pp145-163.